SEQUENCE LISTING

- (1) GENERAL INFORMATION
- (i) APPLICANT: Lal, Preeti Bandman, Olga
- (ii) TITLE OF THE INVENTION: NOVEL HUMAN SODIUM-DEPENDENT PHOSPHATE COTRANSPORTER
- (iii) NUMBER OF SEQUENCES: 7
- (iv) CORRESPONDENCE ADDRESS:
 - (A) ADDRESSEE: Incyte Pharmaceuticals, Inc.
 - (B) STREET: 3174 Porter Drive
 - (C) CITY: Palo Alto
 - (D) STATE: CA
 - (E) COUNTRY: US
 - (F) ZIP: 94304
- (v) COMPUTER READABLE FORM:
 - (A) MEDIUM TYPE: Diskette
 - (B) COMPUTER: IBM Compatible
 - (C) OPERATING SYSTEM: DOS
 - (D) SOFTWARE: FastSEQ Version 2.0
- (vi) CURRENT APPLICATION DATA:
 - (A) APPLICATION NUMBER: To Be Assigned
 - (B) FILING DATE: Filed Herewith
 - (C) CLASSIFICATION:
- (vii) PRIOR APPLICATION DATA:
 - (A) APPLICATION NUMBER:
 - (B) FILING DATE:
- (viii) ATTORNEY/AGENT INFORMATION:
 - (A) NAME: Billings, Lucy J.
 - (B) REGISTRATION NUMBER: 36,749
 - (C) REFERENCE/DOCKET NUMBER: PF-0221 US
- (ix) TELECOMMUNICATION INFORMATION:
 - (A) TELEPHONE: 415-855-0555
 - (B) TELEFAX: 415-845-4166
 - (2) INFORMATION FOR SEQ ID NO:1:
- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 401 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNEȘS: single
 - (D) TOPOLOGY: linear
- (vii) IMMEDIATE SOURCE:
 - (A) LIBRARY: BRAITUT02
 - (B) CLONE: 754412



(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

Met Gln Val Asp Glu Thr Leu Ile Pro Arg Lys Val Pro Ser Leu Cys Ser Ala Arg Tyr Gly Ile Ala Leu Val Leu His Phe Cys Asn Phe Thr 25 Thr Ile Ala Gln Asn Val Ile Met Asn Ile Thr Met Val Ala Met Val 40 Asn Ser Thr Ser Pro Gln Ser Gln Leu Asn Asp Ser Ser Glu Val Leu Pro Val Asp Ser Phe Gly Gly Leu Ser Lys Ala Pro Lys Ser Leu Pro 70 75 Ala Lys Ser Ser Ile Leu Gly Gly Gln Phe Ala Ile Trp Glu Arg Trp 85 90 Gly Pro Pro Gln Glu Arg Ser Arg Leu Cys Ser Ile Ala Leu Ser Gly 105 Met Leu Leu Gly Cys Phe Thr Ala Ile Leu Ile Gly Gly Phe Ile Ser 120 Glu Thr Leu Gly Trp Pro Phe Val Phe Tyr Ile Phe Gly Gly Val Gly 135 Cys Val Cys Cys Leu Leu Trp Phe Val Val Ile Tyr Asp Asp Pro Val 150 155 Ser Tyr Pro Trp Ile Ser Thr Ser Glu Lys Glu Tyr Ile Ile Ser Ser 165 170 Leu Lys Gln Gln Val Gly Ser Ser Lys Gln Pro Leu Pro Ile Lys Ala 180 185 Met Leu Arg Ser Leu Pro Ile Trp Ser Ile Cys Leu Gly Cys Phe Ser 200 205 His Gln Trp Leu Val Ser Thr Met Val Val Tyr Ile Pro Thr Tyr Ile 215 220 Ser Ser Val Tyr His Val Asn Ile Arg Asp Asn Gly Leu Leu Ser Ala 230 235 Leu Pro Phe Ile Val Ala Trp Val Ile Gly Met Val Gly Gly Tyr Leu 250 245 Ala Asp Phe Leu Leu Thr Lys Lys Phe Arg Leu Ile Thr Val Arg Lys 260 265 Ile Ala Thr Ile Leu Gly Ser Leu Pro Ser Ser Ala Leu Ile Val Ser 280 285 Leu Pro Tyr Leu Asn Ser Gly Tyr Ile Thr Ala Thr Ala Leu Leu Thr 295 300 Leu Ser Cys Gly Leu Ser Thr Leu Cys Gln Ser Gly Ile Tyr Ile Asn 315 310 Val Leu Asp Ile Ala Pro Arg Tyr Ser Ser Phe Leu Met Gly Ala Ser 325 330 Arg Gly Phe Ser Ser Ile Ala Pro Val Ile Val Pro Thr Val Ser Gly 340 345 Phe Leu Leu Ser Gln Asp Pro Glu Phe Gly Trp Arg Asn Val Phe Phe 360 Leu Leu Phe Ala Val Asn Leu Leu Gly Leu Leu Phe Tyr Leu Ile Phe 375 380 Gly Glu Ala Asp Val Gln Glu Trp Ala Lys Glu Arg Lys Leu Thr Arg 395 Leu

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1643 base pairs



- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:

- (A) LIBRARY: BRAITUT02
- (B) CLONE: 754412

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

GTCCGGCAAT TCCCGGGTCG ACCCACGCGT CCGCTTGGAG GGACGCTGGG TTCAACTTGA 120 AGCCCTTCCA CAGACATTAA GTCGGTGAAA ACCATTCACT AGGACGAGGAG AAACACAATG 180 GCCACCAAGA CAGAGTTGAG TCCCACAGCA AGGGAGAGCA AGAACCACA AGATATGCAA 240 GTGGATGAGA CACTGATCCC CAGGAAAGTT CCAACGTTAT GTTCTGCTCG CTATGGAATA 300 GCCCTCGTCT TACATTTCTG CAATTCACA ACGATAGCAC AAAATGCAA CATGAACCACT 360 ACCATGGTAG CCATGGTCAA CAGCACAAGC CCTCAATCCC AGCTCAATGA TCCTCTGAG 420 GTGGTGCTG TTGACTCATT TGGTGGCCTA AGTAAGCAC AAAATGTCAT CATGAACATC 360 ACCATGGTAG CCATGGTCAA CAGCACAAGC CCTCAATCCC AGCTCAATGA TTCCTCTGAG 420 GTGCTGCCTG TTGACTCATT TGGTGGCCTA AGTAAAGCCC CAAAGAGTCT TCCTGCAAAG 480 TCCTCAATAC TTGGGGGTCA GTTTGCAATT TGGGAAAGGT GGGCCCTCC ACAAGAACGA 540 AGCAGACTCT GCAGCATTGC TTTATCAGGA ATGTTACTGG GATGCTTTAC TGCCATCCTC 600 ATAGGTGGCT TCATTAGTGA AACCCTTGGG TGGCCCTTTTA TTTTTATATT CTTTTGAAGGT 660 GTTGGCTGTG TCTTCTGCTT TCTCTGGTTT GTTTGTAATAC TCTTTGAAGGT 660 GTTGGCTGTG TCTTCTGCTT TCTCTGGTTT GTTTGTAACAC ACAGGTCGGG 780 TCTTCTAAGC AGCCTCTCC CACAAAGCAT ATGCTCAGAT TCTTCAAACA ACAGGTCGGG 780 TCTTCTAAGC AGCCTCTTCC CATCAAAGCT ATGCTCAGAT TGGTCCATA TTGGTCCATA 840 TGTTTATGTTG CTTGGTGCAT AGGCATGTA GTTAACAAC ACAGGTCGGG 780 TCTTATTGTTG CTTGGTGACA TGATTAACATC AGAGACAAT GACTTCTATC TGCCCTTCCT 960 TTTATTGTTG CTTGGGGTCAT AGGCATGGTG GAGGCCATAC TGGCAGATTT CCTTCCAACC AAAAAGTTTA GACTCATCAC TGTGAGGAAA ATTGCCACAA TGTTTATAT CATCCCACT 1080 TCAACCACTCA TTGTGTCTCT GCCTTACCTC AATTCCACA TTTTAGGAAGA TCTCCCCTCT 1080 TCAACCACTCA TTGTGTCTCT GCCTTTCCTC AATGGCACA TTTTAGGAAGA TCTCCCCTCT 1080 TCAACCACTCA TTGTGTCTCT GCCTTTCCTC ATTCCAGCA TTTTAGGAAGA TCTCCCCTCT 1080 TCAACGACTCA TTGTGCCAC TGTCAGCGA TTTCTTCTTA GTCAGCACA TTTTAGGAAGA TCTCCCCTCT 1140 CTGACGCTCT TTTGCGGATT AAGCACATTC GTCAGTCAC AATTCCAGC TCTCTCCT 1200 GAACTGCTC TTTGCGGATT AAGCACATTC GTCAGTCAC TCTCTCTTA CTTCCAGCCAT TTGCAGCACA TTTTTTTTTT	AGAACGGTGA	GGATGACCGA	CGTATAGGCG	AGAGCCTAGG	TACGCCATGC	CAGGTCACCG	60
GCCACCAAGA CAGAGTTGAG TCCCACAGCA AGGGAGAGCA AGAACGCACA AGATATGCAA GTGGATGAGA CACTGATCCC CAGGAAAGTT CCAAGTTTAT GTTCTGCTCG CTATGGAATA GCCCTCGTCT TACATTTCTG CAATTTCACA ACGATAGCAC AAAATGTCAT CATGAACATC ACCATGGTAG CCATGGTCAA CAGCACAAGC CCTCAATCCC AGCTCAATGA TTCCTCTCAG GTGCTGCCTG TTGACTCATT TGGTGGGCTA AGTAAAGCCC CAAAGAGGTCT TCCTGCAAAG TCCTCAATAC TTGGGGGTCA GTTTGCAATT TGGGAAAGT GGGGCCCTCC ACAAGAACGA AGCAGACTCT GCAGCATTGC TTTATCAGGA ATGTTACTGG GATGCTTTAC TGCCATCCTC AGCAGACTCT GCAGCATTGC TTTATCAGGA ATGTTACTGG GATGCTTTAC TGCCATCCTC ATAGGTGGCT TCATTAGTGA AACCCTTGGG TGGCCCTTTG TCTTCTATAT CTTTGGAGGT 660 GTTGGCTGTG TCTGCTGCT TCTCTGGTTT GTTGTGATTT ATGATGACCC CGTTTCCTAT CCATGGATAA GCACCTCAGA AAAAGAATAC ATCATATCCT CCTTGAAACA ACAGGTCGGG 780 TCTTCTAAGC AGCCTCTCC CATCAAAGCT ATGCTCAGAT CTCTACCCAT TTGGTCCATA 840 TGTTTAGGCT GTTTCAGCCA TCAATAGCTTA GTTAGCACAA TGGTTGTATA CATACCAACT 900 TACATCAGCT CTTGTGACCA TCAATAGCTA GGAGACAATG GACTTCTATC TGCCCTTCCT TTATTTGTTG CCTGGGTCAT AGGCATGGTG GGAGGCTATC TGGCAGATTT CCTTCTAACC AAAAAGTTTA GACTCATCAC TGTGAGGAAA ATTGCCACAA TTGTTCATAC TCCCTTCCT 1080 TCAGCACTCA TTGTGTCCT GCCTTACCTC AATTCCGCT ATTTAGGAAG TCTCCCCTCT 1080 TCAGCACTCA TTGTGTCCT GCCTTACCTC AATTCCGCT ATTTAGGAAG TCTCCCCTCT 1080 GATATTGCT CAAGGTATT CAGCAATTG TGTCAGCAA TTTTAGGAAG TCTCCCCTCT 1080 GCACCTGTCA TTGTCGGGAT AAGCACATTG TGTCAGCAA TTTTAGGAAG TCTCCCCCTCT 1080 GATATTGCT CAAGGTATC CAGTTTCCT AAGCACAA TGTCAGCACA TTTCAGCAC AACTGCCTTG 1140 CTGACGCTCT CTTGCGGATT AAGCACATTG TGTCAGCACA TTTTAGGAAG TCTCCCCCTCT 1080 GATATTGCT CAAGGTATTC CAGTTTTCTC ATGGGAGAAT CAAGAGGATT TCCAGCATA 1260 GCACCTGTCA TTGTCCCAC TGTCAGCGGA TTTCTTCTTA GTCAGCACC TGAGTTTGGG 1320 TGGAGGAATG TCTTCTTCTT GCTGTTTGCC GTTAACCTC TAGGACACC TGAGTTTGGG 1320 TGGAGGAATG TCTTCTTCTT GCTGTTTTCC ATGAGAGAG GAAAACTCAC CTTTCTACCTC 1380 ATATTTGGAG AAGCAGATG CCAAGAATGG GCTAAAAGAG GAAAACTCAC CTTTCTACCTC 1380 ATATTGGAG AAGCAGATG CCAAGAATGG GCTAAAAGAG GAAAACTCAC CTTTCTACCTC 1380 CCGTGATGAA AATACCAGTG AAAAGATTTT TTTTTCCTGT GGCTCTTTTC AATTATGAGA 1560 CCGTGATGAA AATACCAGTG AAAAGATTTT TTTTTCCTGT GGCTCTTTTC AA	GTCCGGCAAT	TCCCGGGTCG	ACCCACGCGT	CCGCTTGGAG	GGACGCTGGG	TTCAACTTGA	120
GTGGATGAGA CACTGATCCC CAGGAAAGTT CCAAGTTTAT GTTCTGCTCG CTATGGAATA 300 GCCCTCGTCT TACATTCTG CAATTTCACA ACGATAGCAC AAAATGTCAT CATGAACATC 360 ACCATGGTAG CCATGGTCAA CAGCACAAGC CCTCAATCCC AGCTCAATGA TTCCTCTGAG 420 GTGCTGCTG TTGACTCATT TGGTGGCCTA AGTAAAGCCC CAAAGAGTCT TCCTGCAAAG 480 TCCTCAATCC TTGGGGGTCA GTTTGCAATT TGGGAAAGGT GGGGCCCTCC ACAAGAACGA 540 AGCAGACTCT GCAGCATTGC TTTATCAGGA ATGTTACTGG GATGCTTTAC TGCCATCCTC 600 ATAGGTGGCT TCATTAGTGA AACCCTTGGG TGGCCCTTTG TCTTCTATAT CTTTGGAGGT 660 GTTGGCTGTG TCTGCTGCA AAAAGAATAC ATCATATCCT CCTTGAAACA ACAGGTCGGG 780 TCTTCTAAGCA AACCCTCGGT TGTTGTGATTT ATGATGACC CGTTTCCTAT 720 CCATGGATAA GCACCTCAGA AAAAGAATAC ATCATATCCT CCTTGAAACA ACAGGTCGGG 780 TCTTCTAAGC AGCCCTCCAAAAGAATAC ATCATATCCT CCTTGAAACA ACAGGTCGGG 780 TCTTCTAAGC AGCCCTCAGA AAAAGAATAC ATCATATCCT CCTTGAAACA ACAGGTCGGG 780 TCTTTAAGCA AGCCTCAGA AAAAGAATAC ATCATATCCT CCTTGAAACA ACAGGTCGGG 780 TACATCAGCT CTGTGTACCA TCTAAAGCT AGGACAATA GACTCTATC TGCCCTTCCT 960 TACATCAGCT CTGTGTACCA TGTTAACATC AGAGACAAT GACTTCTATC TGCCCTTCCT 960 TACATCAGCT CTGTGACCA TGTTAACATC AGAGACAATA GACTCTATC TCCCCTCT 1080 TCAAAAGGTTTA GACTCATCAC TGTGAGGAAA ATTGCCACAA TTTTAGGAAG TCTCCCCTCT 1080 TCAAGCACTCA TTGTGTCTCT GCCTTTCCC AATTCCCACA TTTTAGGAAG TCTCCCCTCT 1080 TCAGCACTCA TTGTGTCTCT GCCTTTCCC AATTCCGCT ATATCACAGC AACTGCCTTG 1140 TCAGCACTCA TTGTGCCATTCC CAGTTTTCCC AATTCCCGCT ATATCACAGC AACTGCCTTG 1140 TCAGCACTCA TTGTACCCA TGTCAGCGAA TTCTTCTTCTA GCCATTCTCT TCAGCCACA TTTTAGGAAG TCTCCCCTCT 1080 GATATTGCC CAAGGTATT CAGTTTTCC ATTCAGCACA TCTCTTCTCT	AGCCCTTCCA	CAGACATTAA	GTCGGTGAAA	ACCATTCACT	AGGAGAGGAG	AAACACAATG	180
GCCCTCGTCT TACATTTCTG CAATTTCACA ACGATAGCAC AAAATGTCAT CATGAACATC 360 ACCATGGTAG CCATGGTCAA CAGCACAAGC CCTCAATCC AGCTCAATGA TTCCTCTGAG 420 GTGCTGCCTG TTGACTCATT TGGTGGCCTA AGTAAAGCCC CAAAGAGTCT TCCTGCAAAG 480 TCCTCAATAC TTGGGGGTCA GTTTGCAATT TGGGAAAGGT GGGGCCCTCC ACAAGAACGA 540 AGCAGACTCT GCAGCATTGC TTTATCAGGA ATGTTACTGG GATGCTTTAC TGCCATCCTC 600 ATAGGTGGCT TCATTAGTGA AACCCTTGGG TGTGCCTTTTG TCTTCATATA CTTTGGAGGT 660 GTTGGCTGTG TCTTGTGCTC TCTCTGGTTT GTTGTGATTT ATGATGACCA CGTTTCCTAT 720 CCATGGATAA GCACCTCAGA AAAAGAATAC ATCATATCCT CCTTGAAACA ACAGGTCGGG 780 TCTTCTAAGC AGCCTCTCC CATCAAAGCT ATGCTCAGAT CTCTTACCCAT TTGGTCCATA 440 TGTTTAGGCT GTTTCAGCA TGATACACAC TGTTACACAC TGTTACCATA AGGACACAC TGGTCTTATA CATACCAACT 900 TACATCAGCT CTGGGTCAT AGGCACACA TGGTCTATA CATACCAACT 900 TACATCAGCT CTGGGTCAT AGGCATGAA ATGCACAAA TGGTTCTATA CATACCAACT 900 TACATCAGCT CTGGGTCAT AGGCATGAA ATTGCCACAA TGGTTCTATA CATACCAACT 900 TACATCAGCT CTGGGTCAT AGGCATGAA ATTGCCACAA TTTTAGGAAG TCCCCTTCCT 960 TCAAAAGTTTA GACTCATCAC TGTGAGGAAA ATTGCCACAA TTTTAGGAAG TCCCCTTCCT 1080 TCAGCACTCA TTGTGTCCAC TGTGAGGAAA ATTGCCACAA TTTTAGGAAG TCCCCCTCT 1080 TCAGCACTCA TTGTGTCCTC GCCTTACCC AATTCCGGCT ATATCACAGC AACTGCCTTG 1140 CTGACCACTC TTGTGTCACT CAGTTTTCTC AATTCCAGC AACTGCCTTG 1140 CTGACCACTC TTGTGACCAC TGTCAGCGA TTTCTTCTTA CAATGCCACA TTCTACCC TGTGAGGAAA ATTGCCACAA TTTTAGGAAG TCTCCCCTCT 1080 GCACCTGTCA TTGTACCAC TGTCAGCGGA TTTCTTCTTA GTCAGGACC TGAGTTTTGG 1320 TGGAGGAATG TCTCTCTTCTT GCTGTTTCTC CAGTTTTCTC ATGGGACCT TTGGAGCATA TCTCAGCCTC TAGGTTTTGG 1320 TGGAGGAATG TCTTCTTCTT GCTGTTTCC GTTAACCTG TAGGACCAC TGGGTTTTGG 1320 TGGAGGAATG TCTTCTTCTT GCTGTTTTCC GTTAACCAG AAAAGCATTT TTGCAGCACA TTGCTTCACCTC TAGGACCA TTGCTTCACCTC TAGGACCA TTGCTTCACCTC TAGGACCA TTGCTTCACCTC TAGGACCA TTGCTTCACCTC TAGGACCA TTGCTTTATAG 1440 AGTTATCCCA CCTTGGATGG AAAAGTCATT TTTTTCTTTT TGAGAGAAA TAGAAACTCA TCGTTTATAGA 1560 TCCCCTTTTTTT TTTTTTTTTTTTTTTTTTTTTTTT	GCCACCAAGA	CAGAGTTGAG	TCCCACAGCA	AGGGAGAGCA	AGAACGCACA	AGATATGCAA	240
ACCATGGTAG CCATGGTCAA CAGCACAAGC CCTCAATCCC AGCTCAATGA TTCCTCTGAG 420 GTGCTGCCTG TTGACTCATT TGGTGGCCTA AGTAAAGCCC CAAAGAGTCT TCCTGCAAAG 480 TCCTCAATAC TTGGGGGTCA GTTTGCAATT TGGGAAAGGT GGGGCCCTCC ACAAGAACGA 540 AGCAGACTCT GCAGCATTGC TTTATCAGGA ATGTTACTGG GATGCTTTAC TGCCATCCTC 600 ATAGGTGGCT TCATTAGTGA AACCCTTGGG TGGCCCTTTG TCTTCTATAT CTTTGGAGGT 660 GTTGGCTGTG TCTGCTGCAT TCTCTGGTTT GTTGTGATTT ATGATGACCC CGTTTCCTAT 720 CCATGGATAA GCACCTCAGA AAAAGAATAC ATCATATCCT CCTTGAAACA ACAGGTCGGG 780 TCTTCTAAGC AGCCTCTCC CATCAAAGCT ATGCTCAGAT CTCTACCCAT TTGGTCCATA 840 TGTTTAGGCT GTTTCAGCCA TCAATGGTTA GTTAGCACAA TGGTTGTATA CATACCAACT 900 TACATCAGCT CTGTGTACCA TGATGACAAC AGGGTCGGG GAGGCTATC TGGCAGATTT CCTTCCTCT 960 TTTATTGTTG CCTTGGTACA AGGAATGGT GGAGGCTATC TGGCAGATTT CCTTCTAACC 1020 AAAAAGTTTA GACTCACAC TGTGAAGAAA ATTGCCACAA TTTTAGGAAG TCTCCCCTCT 1080 TCAGCACTCA TTGTGTCCTC TGCCCTTCCT AATTCCGGCT TTGTGTCATC TGCCCTTCCT AACAACT TGGCAGATT CTTCAGCACA TTTTAGGAAG TCTCCCCTCT 1080 GAAAAAGTTTA GACTCATCAC TGTGAAGAAA ATTGCCACAA TTTTAGGAAG TCTCCCCTCT 1080 GATATTGCCC CTTGCGGATT AAGCACATTG TGTCAGTCAG GGATTTATAT CAATGCCTT 1140 GAAAAGTTTA CAAGGCATC CAAGGAATTC CAGCACATTG TGTCAGTCAG GAATTTATAT CAATGCTTTA 1200 GATATTGCCC CAAGGTATTC CAGCTTTCCT AATTCCGGCT ATTCAGACA TTCGAGCATA 1260 GCACCTGTCA TTGTACCCAC TGTCAGCGGA TTTCTTCTTA GTCAGTCAG GAATTTATAT CAATGCTTTA 1200 GATATTGCCC CAAGGTATTC CAGTTTTCCC GTTAACCTC TAGGAGCAC CAAGAGGATT TCCGAGCATA 1260 GCACCTGTCA TTGTACCCAC TGTCAGCGGA TTTCTTCTTA GTCAGTCAG TAGGAGCACC TGAGTTTTGGG 1320 TGGAGGAATG TCTTCTTCTT GCCGTTTTCCC GTTAACCTGT TAGGACTACT CTTCTACCTC 1380 ATATTTGGAA AAAGATTAT AGGCACCGTA TTGCATAAAA TAGAAAGTTT 1500 CCGTGATGAA AAAAGTCAT TTGTACCTA TAGAAGAGA GAAAACTCAC TCGTTTATGA 1440 AGTTATCCCA CCTTGGATGG AAAAGTCATT TTTTTCCTGT GGCTCTTTTC AATTATGAGA 1560 CCGTGATGAAAAAATTCA CAAGAGGAAT TAAAAATTCA 1500 CCGTGATTGAAAAA TAAAAATTCA GACTTTTTTT TTTTTCCTGT GGCTCTTTTC AATTATGAGA 1560 CCGTGATGAAAAACTCAC TCGTTTATGA 1440 AGTTATCCAT AATTATATAA AAAAATTCA GACTTTTTTTT TTTTTTCCTGT GGCTCTTTTC AATTATAGAA TAAAAATTCA 1560 CCGTGATTCATT ATTTTTTT	${\tt GTGGATGAGA}$	CACTGATCCC	CAGGAAAGTT	CCAAGTTTAT	GTTCTGCTCG	CTATGGAATA	300
GTGCTGCCTG TTGACTCATT TGGTGGCCTA AGTAAAGCCC CAAAGAGTCT TCCTGCAAAG TCCTCAATAC TTGGGGGTCA GTTTGCAATT TGGGAAAGGT GGGGCCCTCC ACAAGAACGA 540 AGCAGACTCT GCAGCATTGC TTTATCAGGA ATGTTACTGG GATGCTTTAC TGCCATCCTC 600 ATAGGTGGCT TCATTAGTGA AACCCTTGGG TGGCCCTTTG TCTTCTATAT CTTTGGAGGT 660 GTTGGCTGTG TCTGCTGCCT TCTCTGGTTT GTTGTGATTT ATGATGACCC CGTTTCCTAT 720 CCATGGATAA GCACCTCAGA AAAAGAATAC ATCATATCCT CCTTGAAACA ACAGGTCGGG 780 TCTTCTAAGC AGCCTCTCC CATCAAAGCT ATGCTCAGAT CTCTACCCAT TTGGTCCATA 840 TGTTTAGGCT GTTTCAGCCA TCAATGGTTA GTTAGCACAA TGGTTGTATA CATACCAACT 900 TACATCAGCT CTGTGTACCA TGTTAACATC AGAGACAATG GACTTCTATC TGCCCTTCCT 960 TTTATTGTTG CCTGGGTCAT AGGCATGGT GGAGGCTATC TGGCAGATTT CCTTCTAACC 1020 AAAAAGTTTA GACTCATCAC TGTGAGGAAA ATTGCCACAA TTTTAGGAGA TCTCCCCTCT 1080 TCAGCACTCA TTGTGTCCT GCCTTACCT AATTCCGGCT ATATCAGAG ACCTCCTCT 1080 GAAAAGTTTA GACTCATCC CAGTTTTCCT AATTCCGGCT ATATCACAGC AACTGCCTTC 1140 GAAAAGTTTA CAAGGCATCA CAGGACAATTG TGTCAGTCAG GAATTTATAT CAATGCTTTA 1200 GAAAAGTTCC CAAGGAATTC CAGTTTTCC AATTCCGGCT ATATCACAGC AACTGCCTTG 1140 GAAAAGTTCC CAAGGAATTC CAGTTTTCCT AGGAGCAAT TTCGAGCATT TCCGACCATA 1260 GAACCTGCTC TTGTACCCC TGTCAGCGGA TTTCTTCTTA GCAGAATGT TTCGAGCATA 1260 GCACCTGTCA TTGTACCCAC TGTCAGCGGA TTTCTTCTTA GTCAGCACAT TTCGAGCATA TTCGAGCATA 1260 GCACCTGTCA TTGTACCCAC TGTCAGCGGA TTTCTTCTTA GTCAGCACA TTTCTACCTC TAGGAGCACT CAAGAGGATT TCCGAGCATA 1260 GCACCTGTCA TTGTACCCAC TGTCAGCGGA TTTCTTCTTA GTCAGCACA TTTCTACCTC TAGGAGCACT CTGTTTTGGG 1320 TGGAGGAATG TCTTCTTCTT GCTGTTTTCC GTTAACCTG TAGGACCC TGAGTTTTGGA 1340 ATATTTGGAG AAAAGTCAT AGGCACCGTA TTGCATAAAA TAGAAGGCTT 1500 CCGTGATGAA AAAAGTCAT TTGTACCAC TCGTTTATGA 1440 AGTTATCCCA CCTTGGATGG AAAAGTCAT TTGTACCAC TCGTTTATGA 1440 AGTTATCCAC CCTTGGATGG AAAAGTCAT TTGTACTAC TCGTTTATGA 1440 AGTTATCCAC CCTTGGATGG AAAAGTCAT TTGTACCTC TGTTTATGA 1440 AGTTATCCAT ATTTTTTT TTTTTCCTGT GGCTCTTTTC AATTATGAGA 1560 CCGTGATGAAAAACTCAC TCGTTTATGA 1440 AGTTATCCAC TTGTGAGCAT TTTTTTTT TTTTTCCTGT GGCTCTTTTC AATTATGAGA 1560 TAAAAAATTCA TAAAAATTCA 1560 TAAAGATGAA TAAAAATTCA 1560 TAAAGATGAA TAAAAATTC	GCCCTCGTCT	\mathbf{T} ACATTTCTG	CAATTTCACA	ACGATAGCAC	AAAATGTCAT	CATGAACATC	360
TCCTCAATAC TTGGGGGTCA GTTTGCAATT TGGGAAAGGT GGGGCCCTCC ACAAGAACGA AGCAGACTCT GCAGCATTGC TTTATCAGGA ATGTTACTGG GATGCTTTAC TGCCATCCTC 600 ATAGGTGGCT TCATTAGTGA AACCCTTGGG TGGCCCTTTG TCTTCTATAT CTTTGGAGGT 660 GTTGGCTGTG TCTGCTGCCT TCTCTGGTTT GTTGTGATTT ATGATGACCC CGTTTCCTAT 720 CCATGGATAA GCACCTCAGA AAAAGAATAC ATCATATCCT CCTTGAAACA ACAGGTCGGG 780 TCTTCTAAGC AGCCTCTCC CATCAAAGCT ATGCTCAGAT CTCTACCCAT TTGGTCCATA 840 TGTTTAGGCT GTTTCAGCCA TCAATGGTTA GTTAGCACAA TGGTTGTATA CATACCAACT 900 TACATCAGCT CTGTGTACCA TGTTAACATC AGAGACAATG GACTTCTATC TGCCCTTCCT 960 TTTATTGTTG CCTGGGTCAT AGGCATGGTG GGAGGCTATC TGGCAGATTT CCTTCTAACC 1020 AAAAAGTTTA GACTCATCAC TGTGAGGAAA ATTGCCACAA TTTTAGGAAG TCTCCCCTCT 1080 TCAGCACTCA TTGTGTCTCT GCCTTACCTC AATTCCGGCT ATATCACAGC AACTGCCTTG 1140 CTGACGCTCT CTTGCGGATT AAGCACATTG TGTCAGTCAG GGATTTATAT CAATGTCTTA 1200 GATATTGCTC CAAGGTATTC CAGTTTTCTC ATGGGAGCAC TCAAGAGGATT TCCGAGCATA 1260 GCACCTGTCA TTGTACCACA TGTCAGCGGA TTTCTTCTTA GTCAGGACC TGAGTTTGGG 1320 TGGAGGAATG TCTCTCTTT GCTGTTTGCC GTTAACCTG TAGGACCC TGAGTTTGGG 1320 TGGAGGAATG TCTCTTCTT GCTGTTTGCC GTTAACCTG TAGGACCC TGAGTTTGGG 1320 TGGAGGAATG TCTCTTCTTT GCTGTTTGCC GTTAACCTG TAGGACCAC TCGTTTATGA 1440 AGTTATCCCA CCTTGGATGG AAAAGTCATT AGGCACCGTA TTGCATAAAA TAGAAGGCTT 1500 CCGTGATGAA AATACCAGTG AAAAGATTTT TTTTTCCTGT GGCTCTTTTC AATTATGAGA 1560 TCAGTTCATT ATTTTATTCA GACTTTTTTT TGAGAGAAAT GTAAGATGAA TAAAAATTCA 1620	ACCATGGTAG	CCATGGTCAA	CAGCACAAGC	CCTCAATCCC	AGCTCAATGA	TTCCTCTGAG	420
AGCAGACTCT GCAGCATTGC TTTATCAGGA ATGTTACTGG GATGCTTTAC TGCCATCCTC 600 ATAGGTGGCT TCATTAGTGA AACCCTTGGG TGGCCCTTTG TCTTCTATAT CTTTGGAGGT 660 GTTGGCTGTG TCTGCTGCCT TCTCTGGTTT GTTGTGATTT ATGATGACCC CGTTTCCTAT 720 CCATGGATAA GCACCTCAGA AAAAGAATAC ATCATATCCT CCTTGAAACA ACAGGTCGGG 780 TCTTCTAAGC AGCCTCTCC CATCAAAGCT ATGCTCAGAT CTCTACCCAT TTGGTCCATA 840 TGTTTAGGCT GTTTCAGCCA TCAATGGTTA GTTAGCACAA TGGTTGTATA CATACCAACT 900 TACATCAGCT CTGTGTACCA TGTTAACATC AGAGACAATG GACTTCTATC TGCCCTTCCT 960 TTTATTGTTG CCTGGGTCAT AGGCATGGTG GGAGGCTATC TGGCAGATTT CCTTCTAACC 1020 AAAAAGTTTA GACTCATCAC TGTGAGGAAA ATTGCCACAA TTTTAGGAAG TCTCCCCTCT 1080 TCAGCACTCA TTGTGTCTCT GCCTTACCTC AATTCCGGCT ATATCACAGC AACTGCCTTG 1140 CTGACGCTCT CTTGCGGATT AAGCACATTG TGTCAGTCAG GGATTTTATAT CAATGTCTTA 1200 GATATTGCTC CAAGGTATTC CAGTTTTCTC ATGGGAGCAT CAAGAGGATT TCCGAGCATA 1260 GCACCTGTCA TTGTACCCAC TGTCAGCGGA TTTCTTCTTA GTCAGGACCC TGAGTTTGGG 1320 TGGAGGAATG TCTTCTTCTT GCTGTTTGCC GTTAACCTG TAGGACTACT CTTCTACCTC 1380 ATATTTGGAG AAGCAGATGT CCAAGAATGG GCTAAAGAGA GAAAACTCAC TCGTTTATGA 1440 AGTTATCCCA CCTTGGATGG AAAAGTCATT AGGCACCGTA TTGCATAAAA TAGAAGGCTT 1500 CCGTGATGAA AATACCAGTG AAAAGATTTT TTTTTCCTGT GGCTCTTTTC AATTATGAGA 1560 TCAGGTTCATT ATTTTATTCA GACCTTTTTT TGAGGAAAAT GTAAGATGAA TAAAAAATTCA 1620	GTGCTGCCTG	TTGACTCATT	TGGTGGCCTA	AGTAAAGCCC	CAAAGAGTCT	TCCTGCAAAG	480
ATAGGTGGCT TCATTAGTGA AACCCTTGGG TGGCCCTTTG TCTTCTATAT CTTTGGAGGT 660 GTTGGCTGTG TCTGCTGCCT TCTCTGGTTT GTTGTGATTT ATGATGACCC CGTTTCCTAT 720 CCATGGATAA GCACCTCAGA AAAAGAATAC ATCATATCCT CCTTGAAACA ACAGGTCGGG 780 TCTTCTAAGC AGCCTCTCC CATCAAAGCT ATGCTCAGAT CTCTACCCAT TTGGTCCATA 840 TGTTTAGGCT GTTTCAGCCA TCAATGGTTA GTTAGCACAA TGGTTGTATA CATACCAACT 900 TACATCAGCT CTGTGTACCA TGTTAACATC AGAGACAATG GACTTCTATC TGCCCTTCCT 960 TTTATTGTTG CCTGGGTCAT AGGCATGGTG GGAGGCTATC TGGCAGATTT CCTTCTAACC 1020 AAAAAGTTTA GACTCATCAC TGTGAGGAAA ATTGCCACAA TTTTAGGAAG TCTCCCCTCT 1080 TCAGCACTCA TTGTGTCTCT GCCTTACCTC AATTCCGGCT ATATCACAGC AACTGCCTTG 1140 CTGACGCTCT CTTGCGGATT AAGCACATTG TGTCAGTCAG GGATTTATAT CAATGTCTTA 1200 GATATTGCTC CAAGGTATC CAGTTTTCTC ATGGGAGCAT CAAGAGGATT TCCGAGCATA 1260 GCACCTGTCA TTGTACCCAC TGTCAGCGGA TTTCTTCTTA GTCAGGACCC TGAGTTTGGG 1320 TGGAGGAATG TCTTCTTCTT GCTGTTTGCC GTTAACCTGT TAGGACTACT CTTCTACCTC 1380 ATATTTGGAG AAGCAGATGT CCAAGAATAG GCTAAAGAGA GAAAACTCAC TCGTTTATGA 1440 AGTTATCCCA CCTTGGATGG AAAAGTCATT AGGCACCGTA TTGCATAAAA TAGAAGGCTT 1500 CCGTGATGAA AATACCAGTG AAAAGATTTT TTTTTCCTGT GGCTCTTTTC AATTATGAGA 1560 TCAGGTTCATT ATTTTATTCA GACTTTTTTT TGAGAGAAAT GTAAGATGAA TAAAAATTCA 1620	TCCTCAATAC	TTGGGGGTCA	GTTTGCAATT	TGGGAAAGGT	GGGGCCCTCC	ACAAGAACGA	540
GTTGGCTGTG TCTGCTGCT TCTCTGGTTT GTTGTGATTT ATGATGACCC CGTTTCCTAT 720 CCATGGATAA GCACCTCAGA AAAAGAATAC ATCATATCCT CCTTGAAACA ACAGGTCGGG 780 TCTTCTAAGC AGCCTCTTCC CATCAAAGCT ATGCTCAGAT CTCTACCCAT TTGGTCCATA 840 TGTTTAGGCT GTTTCAGCCA TCAATGGTTA GTTAGCACAA TGGTTGTATA CATACCAACT 900 TACATCAGCT CTGTGTACCA TGTTAACATC AGAGACAATG GACTTCTATC TGCCCTTCCT 960 TTTATTGTTG CCTGGGTCAT AGGCATGGTG GGAGGCTATC TGGCAGATTT CCTTCTAACC 1020 AAAAAGTTTA GACTCATCAC TGTGAGGAAA ATTGCCACAA TTTTAGGAAG TCTCCCCTCT 1080 TCAGCACTCA TTGTGTCTCT GCCTTACCTC AATTCCGGCT ATATCACAGC AACTGCCTTG 1140 CTGACGCTCT CTTGCGGATT AAGCACATTG TGTCAGTCAG GGATTTATAT CAATGTCTTA 1200 GATATTGCTC CAAGGTATTC CAGTTTTCTC ATGGGAGCAT CAAGAGGATT TTCGAGCATA 1260 GCACCTGTCA TTGTACCCAC TGTCAGCGGA TTTCTTCTTA GTCAGGACCC TGAGTTTGGG 1320 TGGAGGAATG TCTTCTTCTT GCTGTTTGCC GTTAACCTGT TAGGACTACT CTTCTACCTC 1380 ATATTTGGAG AAGCAGATGT CCAAGAATGG GCTAAAGAGA GAAAACTCAC TCGTTTATGA 1440 AGTTATCCCA CCTTGGATGG AAAAGTCATT AGGCACCGTA TTGCATAAAA TAGAAGGCTT 1500 CCGTGATGAA AATACCAGTG AAAAGATTTT TTTTTCCTGT GGCTCTTTTC AATTATGAGA 1560 TCAGTTCATT ATTTTATTCA GACCTTTTTT TGAGAGAAAT GTAAGATGAA TAAAAATTCA 1620	AGCAGACTCT	GCAGCATTGC	TTTATCAGGA	ATGTTACTGG	GATGCTTTAC	TGCCATCCTC	600
CCATGGATAA GCACCTCAGA AAAAGAATAC ATCATATCCT CCTTGAAACA ACAGGTCGGG 780 TCTTCTAAGC AGCCTCTTCC CATCAAAGCT ATGCTCAGAT CTCTACCCAT TTGGTCCATA 840 TGTTTAGGCT GTTTCAGCCA TCAATGGTTA GTTAGCACAA TGGTTGTATA CATACCAACT 900 TACATCAGCT CTGTGTACCA TGTTAACATC AGAGACAATG GACTTCTATC TGCCCTTCCT 960 TTTATTGTTG CCTGGGTCAT AGGCATGGTG GGAGGCTATC TGGCAGATTT CCTTCTAACC 1020 AAAAAGTTTA GACTCATCAC TGTGAGGAAA ATTGCCACAA TTTTAGGAAG TCTCCCCTCT 1080 TCAGCACTCA TTGTGTCTCT GCCTTACCTC AATTCCGGCT ATATCACAGC AACTGCCTTG 1140 CTGACGCTCT CTTGCGGATT AAGCACATTG TGTCAGTCAG GGATTTATAT CAATGTCTTA 1200 GATATTGCTC CAAGGTATTC CAGTTTTCTC ATGGGAGCAT CAAGAGGATT TTCGAGCATA 1260 GCACCTGTCA TTGTACCCAC TGTCAGCGGA TTTCTTCTTA GTCAGGACCC TGAGTTTGGG 1320 TGGAGGAATG TCTTCTTCTT GCTGTTTGCC GTTAACCTGT TAGGACTACT CTTCTACCTC 1380 ATATTTGGAG AAGCAGATGT CCAAGAATGG GCTAAAGAGA GAAAACTCAC TCGTTTATGA 1440 AGTTATCCCA CCTTGGATGG AAAAGTCATT AGGCACCGTA TTGCATAAAAA TAGAAGGCTT 1500 CCGTGATGAA AATACCAGTG AAAAGATTTT TTTTTCCTGT GGCTCTTTTC AATTATGAGA 1560 TCAGTTCATT ATTTTATTCA GACTTTTTT TGAGAGAAAT GTAAGATGAA TAAAAAATTCA 1620	ATAGGTGGCT	TCATTAGTGA	AACCCTTGGG	TGGCCCTTTG	TCTTCTATAT	CTTTGGAGGT	660
TCTTCTAAGC AGCCTCTTCC CATCAAGCT ATGCTCAGAT CTCTACCCAT TTGGTCCATA TGTTTAGGCT GTTTCAGCCA TCAATGGTTA GTTAGCACAA TGGTTGTATA CATACCAACT 900 TACATCAGCT CTGTGTACCA TGTTAACATC AGAGACAATG GACTTCTATC TGCCCTTCCT 960 TTTATTGTTG CCTGGGTCAT AGGCATGGTG GGAGGCTATC TGGCAGATTT CCTTCTAACC 1020 AAAAAGTTTA GACTCATCAC TGTGAGGAAA ATTGCCACAA TTTTAGGAAG TCTCCCCTCT 1080 TCAGCACTCA TTGTGTCTCT GCCTTACCTC AATTCCGGCT ATATCACAGC AACTGCCTTG 1140 CTGACGCTCT CTTGCGGATT AAGCACATTG TGTCAGTCAG GGATTTATAT CAATGTCTTA 1200 GATATTGCTC CAAGGTATTC CAGTTTTCTC ATGGGAGCAT CAAGAGGATT TTCGAGCATA 1260 GCACCTGTCA TTGTACCCAC TGTCAGCGGA TTTCTTCTTA GTCAGGACCC TGAGTTTGGG 1320 TGGAGGAATG TCTTCTTCTT GCTGTTTGCC GTTAACCTGT TAGGACTACT CTTCTACCTC 1380 ATATTTGGAG AAGCAGATGT CCAAGAATGG GCTAAAGAGA GAAAACTCAC TCGTTTATGA 1440 AGTTATCCCA CCTTGGATGG AAAAGTCATT AGGCACCGTA TTGCATAAAA TAGAAGGCTT 1500 CCGTGATGAA AATACCAGTG AAAAGATTTT TTTTTCCTGT GGCTCTTTTC AATTATGAGA 1560 TCAGTTCATT ATTTTATTCA GACTTTTTTT TGAGAGAAAT GTAAGATGAA TAAAAAATTCA 1620	GTTGGCTGTG	TCTGCTGCCT	TCTCTGGTTT	GTTGTGATTT	ATGATGACCC	CGTTTCCTAT	720
TGTTTAGGCT GTTTCAGCCA TCAATGGTTA GTTAGCACAA TGGTTGTATA CATACCAACT 900 TACATCAGCT CTGTGTACCA TGTTAACATC AGAGACAATG GACTTCTATC TGCCCTTCCT 960 TTTATTGTTG CCTGGGTCAT AGGCATGGTG GGAGGCTATC TGGCAGATTT CCTTCTAACC 1020 AAAAAGTTTA GACTCATCAC TGTGAGGAAA ATTGCCACAA TTTTAGGAAG TCTCCCCTCT 1080 TCAGCACTCA TTGTGTCTCT GCCTTACCTC AATTCCGGCT ATATCACAGC AACTGCCTTG 1140 CTGACGCTCT CTTGCGGATT AAGCACATTG TGTCAGTCAG GGATTTATAT CAATGTCTTA 1200 GATATTGCTC CAAGGTATTC CAGTTTTCTC ATGGGAGCAT CAAGAGGATT TTCGAGCATA 1260 GCACCTGTCA TTGTACCCAC TGTCAGCGGA TTTCTTCTTA GTCAGGACCC TGAGTTTGGG 1320 TGGAGGAATG TCTTCTTCTT GCTGTTTGCC GTTAACCTGT TAGGACTACT CTTCTACCTC 1380 ATATTTGGAG AAGCAGATGT CCAAGAATGG GCTAAAGAGA GAAAACTCAC TCGTTTATGA 1440 AGTTATCCCA CCTTGGATGG AAAAGTCATT AGGCACCGTA TTGCATAAAA TAGAAGGCTT 1500 CCGTGATGAA AATACCAGTG AAAAGATTTT TTTTTCCTGT GGCTCTTTTC AATTATGAGA 1560 TCAGTTCATT ATTTTATTCA GACTTTTTTT TGAGAGAAAT GTAAGATGAA TAAAAAATTCA 1620	CCATGGATAA	GCACCTCAGA	AAAAGAATAC	ATCATATCCT	CCTTGAAACA	ACAGGTCGGG	780
TACATCAGCT CTGTGTACCA TGTTAACATC AGAGACAATG GACTTCTATC TGCCCTTCCT TTTATTGTTG CCTGGGTCAT AGGCATGGTG GGAGGCTATC TGGCAGATTT CCTTCTAACC 1020 AAAAAGTTTA GACTCATCAC TGTGAGGAAA ATTGCCACAA TTTTAGGAAG TCTCCCCTCT 1080 TCAGCACTCA TTGTGTCTCT GCCTTACCTC AATTCCGGCT ATATCACAGC AACTGCCTTG 1140 CTGACGCTCT CTTGCGGATT AAGCACATTG TGTCAGTCAG GGATTTATAT CAATGTCTTA 1200 GATATTGCTC CAAGGTATTC CAGTTTTCTC ATGGGAGCAT CAAGAGGATT TTCGAGCATA 1260 GCACCTGTCA TTGTACCCAC TGTCAGCGGA TTTCTTCTTA GTCAGGACCC TGAGTTTGGG 1320 TGGAGGAATG TCTTCTTCTT GCTGTTTGCC GTTAACCTGT TAGGACTACT CTTCTACCTC 1380 ATATTTGGAG AAGCAGATGT CCAAGAATGG GCTAAAGAGA GAAAACTCAC TCGTTTATGA 1440 AGTTATCCCA CCTTGGATGG AAAAGTCATT AGGCACCGTA TTGCATAAAA TAGAAGGCTT 1500 CCGTGATGAA AATACCAGTG AAAAGATTTT TTTTTCCTGT GGCTCTTTTC AATTATGAGA 1560 TCAGTTCATT ATTTTATTCA GACTTTTTTT TGAGAGAAAT GTAAGATGAA TAAAAAATTCA 1620	TCTTCTAAGC	AGCCTCTTCC	CATCAAAGCT	ATGCTCAGAT	CTCTACCCAT	TTGGTCCATA	840
TTTATTGTTG CCTGGGTCAT AGGCATGGTG GGAGGCTATC TGGCAGATTT CCTTCTAACC AAAAAAGTTTA GACTCATCAC TGTGAGGAAA ATTGCCACAA TTTTAGGAAG TCTCCCCTCT 1080 TCAGCACTCA TTGTGTCTCT GCCTTACCTC AATTCCGGCT ATATCACAGC AACTGCCTTG 1140 CTGACGCTCT CTTGCGGATT AAGCACATTG TGTCAGTCAG GGATTTATAT CAATGTCTTA 1200 GATATTGCTC CAAGGTATTC CAGTTTTCTC ATGGGAGCAT CAAGAGGATT TTCGAGCATA 1260 GCACCTGTCA TTGTACCCAC TGTCAGCGGA TTTCTTCTTA GTCAGGACCC TGAGTTTGGG 1320 TGGAGGAATG TCTTCTTCTT GCTGTTTGCC GTTAACCTGT TAGGACTACT CTTCTACCTC 1380 ATATTTGGAG AAGCAGATGT CCAAGAATGG GCTAAAGAGA GAAAACTCAC TCGTTTATGA 1440 AGTTATCCCA CCTTGGATGG AAAAGTCATT AGGCACCGTA TTGCATAAAA TAGAAGGCTT 1500 CCGTGATGAA AATACCAGTG AAAAGATTTT TTTTTCCTGT GGCTCTTTTC AATTATGAGA 1560 TCAGTTCATT ATTTTATTCA GACTTTTTT TGAGAGAAAT GTAAGATGAA TAAAAAATTCA 1620	TGTTTAGGCT	GTTTCAGCCA	TCAATGGTTA	GTTAGCACAA	${\tt TGGTTGTATA}$	CATACCAACT	900
AAAAAGTTTA GACTCATCAC TGTGAGGAAA ATTGCCACAA TTTTAGGAAG TCTCCCCTCT 1080 TCAGCACTCA TTGTGTCTCT GCCTTACCTC AATTCCGGCT ATATCACAGC AACTGCCTTG 1140 CTGACGCTCT CTTGCGGATT AAGCACATTG TGTCAGTCAG GGATTTATAT CAATGTCTTA 1200 GATATTGCTC CAAGGTATTC CAGTTTTCTC ATGGGAGCAT CAAGAGGATT TTCGAGCATA 1260 GCACCTGTCA TTGTACCCAC TGTCAGCGGA TTTCTTCTTA GTCAGGACCC TGAGTTTGGG 1320 TGGAGGAATG TCTTCTTCTT GCTGTTTGCC GTTAACCTGT TAGGACTACT CTTCTACCTC 1380 ATATTTGGAG AAGCAGATGT CCAAGAATGG GCTAAAGAGA GAAAACTCAC TCGTTTATGA 1440 AGTTATCCCA CCTTGGATGG AAAAGTCATT AGGCACCGTA TTGCATAAAA TAGAAGGCTT 1500 CCGTGATGAA AATACCAGTG AAAAGATTTT TTTTTCCTGT GGCTCTTTTC AATTATGAGA 1560 TCAGTTCATT ATTTTATTCA GACTTTTTT TGAGAGAAAT GTAAGATGAA TAAAAAATTCA 1620	TACATCAGCT	CTGTGTACCA	TGTTAACATC	AGAGACAATG	GACTTCTATC	TGCCCTTCCT	960
TCAGCACTCA TTGTGTCTCT GCCTTACCTC AATTCCGGCT ATATCACAGC AACTGCCTTG 1140 CTGACGCTCT CTTGCGGATT AAGCACATTG TGTCAGTCAG GGATTTATAT CAATGTCTTA 1200 GATATTGCTC CAAGGTATTC CAGTTTTCTC ATGGGAGCAT CAAGAGGATT TTCGAGCATA 1260 GCACCTGTCA TTGTACCCAC TGTCAGCGGA TTTCTTCTTA GTCAGGACCC TGAGTTTGGG 1320 TGGAGGAATG TCTTCTTCTT GCTGTTTGCC GTTAACCTGT TAGGACTACT CTTCTACCTC 1380 ATATTTGGAG AAGCAGATGT CCAAGAATGG GCTAAAGAGA GAAAACTCAC TCGTTTATGA 1440 AGTTATCCCA CCTTGGATGG AAAAGTCATT AGGCACCGTA TTGCATAAAA TAGAAGGCTT 1500 CCGTGATGAA AATACCAGTG AAAAGATTTT TTTTTCCTGT GGCTCTTTTC AATTATGAGA 1560 TCAGTTCATT ATTTTATTCA GACTTTTTTT TGAGAGGAAAT GTAAGATGAA TAAAAAATTCA 1620	TTTATTGTTG	CCTGGGTCAT	AGGCATGGTG	GGAGGCTATC	TGGCAGATTT	CCTTCTAACC	1020
CTGACGCTCT CTTGCGGATT AAGCACATTG TGTCAGTCAG GGATTTATAT CAATGTCTTA 1200 GATATTGCTC CAAGGTATTC CAGTTTTCTC ATGGGAGCAT CAAGAGGATT TTCGAGCATA 1260 GCACCTGTCA TTGTACCCAC TGTCAGCGGA TTTCTTCTTA GTCAGGACCC TGAGTTTGGG 1320 TGGAGGAATG TCTTCTTCTT GCTGTTTGCC GTTAACCTGT TAGGACTACT CTTCTACCTC 1380 ATATTTGGAG AAGCAGATGT CCAAGAATGG GCTAAAGAGA GAAAACTCAC TCGTTTATGA 1440 AGTTATCCCA CCTTGGATGG AAAAGTCATT AGGCACCGTA TTGCATAAAA TAGAAGGCTT 1500 CCGTGATGAA AATACCAGTG AAAAGATTTT TTTTTCCTGT GGCTCTTTTC AATTATGAGA 1560 TCAGTTCATT ATTTTATTCA GACTTTTTT TGAGAGAAAT GTAAGATGAA TAAAAAATTCA 1620	AAAAAGTTTA	GACTCATCAC	TGTGAGGAAA	ATTGCCACAA	TTTTAGGAAG	TCTCCCCTCT	1080
GATATTGCTC CAAGGTATTC CAGTTTTCTC ATGGGAGCAT CAAGAGGATT TTCGAGCATA 1260 GCACCTGTCA TTGTACCCAC TGTCAGCGGA TTTCTTCTTA GTCAGGACCC TGAGTTTGGG 1320 TGGAGGAATG TCTTCTTCTT GCTGTTTGCC GTTAACCTGT TAGGACTACT CTTCTACCTC 1380 ATATTTGGAG AAGCAGATGT CCAAGAATGG GCTAAAGAGA GAAAACTCAC TCGTTTATGA 1440 AGTTATCCCA CCTTGGATGG AAAAGTCATT AGGCACCGTA TTGCATAAAA TAGAAGGCTT 1500 CCGTGATGAA AATACCAGTG AAAAGATTTT TTTTTCCTGT GGCTCTTTTC AATTATGAGA 1560 TCAGTTCATT ATTTTATTCA GACTTTTTT TGAGAGAAAT GTAAGATGAA TAAAAAATTCA 1620	TCAGCACTCA	TTGTGTCTCT	GCCTTACCTC	AATTCCGGCT	ATATCACAGC	AACTGCCTTG	1140
GCACCTGTCA TTGTACCCAC TGTCAGCGGA TTTCTTCTTA GTCAGGACCC TGAGTTTGGG 1320 TGGAGGAATG TCTTCTTCTT GCTGTTTGCC GTTAACCTGT TAGGACTACT CTTCTACCTC 1380 ATATTTGGAG AAGCAGATGT CCAAGAATGG GCTAAAGAGA GAAAACTCAC TCGTTTATGA 1440 AGTTATCCCA CCTTGGATGG AAAAGTCATT AGGCACCGTA TTGCATAAAA TAGAAGGCTT 1500 CCGTGATGAA AATACCAGTG AAAAGATTTT TTTTTCCTGT GGCTCTTTTC AATTATGAGA 1560 TCAGTTCATT ATTTTATTCA GACTTTTTT TGAGAGAAAT GTAAGATGAA TAAAAATTCA 1620	CTGACGCTCT	CTTGCGGATT	AAGCACATTG	TGTCAGTCAG	GGATTTATAT	CAATGTCTTA	1200
TGGAGGAATG TCTTCTTCTT GCTGTTTGCC GTTAACCTGT TAGGACTACT CTTCTACCTC 1380 ATATTTGGAG AAGCAGATGT CCAAGAATGG GCTAAAGAGA GAAAACTCAC TCGTTTATGA 1440 AGTTATCCCA CCTTGGATGG AAAAGTCATT AGGCACCGTA TTGCATAAAA TAGAAGGCTT 1500 CCGTGATGAA AATACCAGTG AAAAGATTTT TTTTTCCTGT GGCTCTTTTC AATTATGAGA 1560 TCAGTTCATT ATTTTATTCA GACTTTTTT TGAGAGAAAT GTAAGATGAA TAAAAAATTCA 1620	GATATTGCTC	CAAGGTATTC	CAGTTTTCTC	ATGGGAGCAT	CAAGAGGATT	TTCGAGCATA	1260
ATATTTGGAG AAGCAGATGT CCAAGAATGG GCTAAAGAGA GAAAACTCAC TCGTTTATGA 1440 AGTTATCCCA CCTTGGATGG AAAAGTCATT AGGCACCGTA TTGCATAAAA TAGAAGGCTT 1500 CCGTGATGAA AATACCAGTG AAAAGATTTT TTTTTCCTGT GGCTCTTTTC AATTATGAGA 1560 TCAGTTCATT ATTTTATTCA GACTTTTTT TGAGAGAAAT GTAAGATGAA TAAAAAATTCA 1620	GCACCTGTCA	TTGTACCCAC	TGTCAGCGGA	TTTCTTCTTA	GTCAGGACCC	TGAGTTTGGG	1320
AGTTATCCCA CCTTGGATGG AAAAGTCATT AGGCACCGTA TTGCATAAAA TAGAAGGCTT 1500 CCGTGATGAA AATACCAGTG AAAAGATTTT TTTTTCCTGT GGCTCTTTTC AATTATGAGA 1560 TCAGTTCATT ATTTTATTCA GACTTTTTT TGAGAGAAAT GTAAGATGAA TAAAAAATTCA 1620	TGGAGGAATG	TCTTCTTCTT	GCTGTTTGCC	GTTAACCTGT	TAGGACTACT	CTTCTACCTC	1380
CCGTGATGAA AATACCAGTG AAAAGATTTT TTTTTCCTGT GGCTCTTTTC AATTATGAGA 1560 TCAGTTCATT ATTTTATTCA GACTTTTTT TGAGAGAAAT GTAAGATGAA TAAAAAATTCA 1620	ATATTTGGAG	AAGCAGATGT	CCAAGAATGG	GCTAAAGAGA	GAAAACTCAC	TCGTTTATGA	1440
TCAGTTCATT ATTTTATTCA GACTTTTTT TGAGAGAAAT GTAAGATGAA TAAAAATTCA 1620	AGTTATCCCA	CCTTGGATGG	AAAAGTCATT	AGGCACCGTA	TTGCATAAAA	TAGAAGGCTT	1500
	CCGTGATGAA	AATACCAGTG	AAAAGATTTT	TTTTTCCTGT	GGCTCTTTTC	AATTATGAGA	1560
AATAAAATGA TAACTAAGAA TGC 1643	TCAGTTCATT	ATTTTATTCA	GACTTTTTTT	TGAGAGAAAT	GTAAGATGAA	TAAAAATTCA	1620
	AATAAAATGA	TAACTAAGAA	TGC				1643

(2) INFORMATION FOR SEQ ID NO:3:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 467 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (vii) IMMEDIATE SOURCE:
 - (A) LIBRARY: GenBank
 - (B) CLONE: 450532

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

 Met Gln Met Asp Asn Arg Leu Pro Pro Lys Lys Val Pro Gly Phe Cys

 1
 5
 10
 15

 Ser Phe Arg Tyr Gly Leu Ser Phe Leu Val His Cys Cys Asn Val Ile
 20
 25
 30

 Ile Thr Ala Gln Arg Ala Cys Leu Asn Leu Thr Met Val Val Met Val
 35
 40
 45

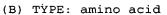


Asn	Ser 50	Thr	Asp	Pro	His	Gly 55	Leu	Pro	Asn	Thr	Ser 60	Thr	Lys	Lys	Leu
Leu 65	Asp	Asn	Ile	Lys	Asn 70	Pro	Met	Tyr	Asn	Trp 75	Ser	Pro	Asp	Ile	Gln 80
Gly	Ile	Ile	Leu	Ser 85	Ser	Thr	Ser	Tyr	Gly 90	Val	Ile	Ile	Ile	G1n 95	Val
Pro	Val	Gly	Tyr 100	Phe	Ser	Gly	Ile	Tyr 105	Ser	Thr	Lys	Lys	Met 110	Ile	Gly
Phe	Ala	Leu 115	Cys	Leu	Ser	Ser	Val 120	Leu	Ser	Leu	Leu	Ile 125	Pro	Pro	Ala
	130					135					140			Gln	
145					150					155				Val	160
				165					170					Thr 175	
			180					185					190	Val	
		195					200					205		Ala	
_	210			_		215	_				220	_		Asp	
225	_			_	230					235		_		Thr	240
				245					250					Ile 255	
			260					265					270	Ser	
		275					280					285		Met	
	290					295					300			Leu	
Ser 305	Leu	Pro	Tyr	Leu	Phe 310	Ala	Trp	Ile	Cys	Gly 315	Asn	Leu	Ala	Gly	Gln 320
		_		325			_		330					Ala 335	
			340					345					350	Phe	
		355					360					365		Ile	
	370					375					380			Val	
385					390					395				Lys	400
				405					410					Thr 415	
			420					425					430	Lys	
		435					440					445		Tyr	
Ile	Val. 450	Ala	Thr	Ala	Glu	Ile 455	Gln	Asp	Trp	Ala	Lys 460	Glu	Lys	Gln	His
Thr 465	Arg	Leu													

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 560 amino acids



- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:

- (A) LIBRARY: GenBank
- (B) CLONE: 507415

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

Met Glu Phe Arg Gln Glu Glu Phe Arg Lys Leu Ala Gly Arg Ala Leu Gly Arg Leu His Arg Leu Leu Glu Lys Arg Gln Glu Gly Ala Glu Thr 20 25 Leu Glu Leu Ser Ala Asp Gly Arg Pro Val Thr Thr His Thr Arg Asp 40 Pro Pro Val Val Asp Cys Thr Cys Phe Gly Leu Pro Arg Arg Tyr Ile Ile Ala Ile Met Ser Gly Leu Gly Phe Cys Ile Ser Phe Gly Ile Arg 70 75 Cys Asn Leu Gly Val Ala Ile Val Ser Met Val Asn Asn Ser Thr Thr 90 His Arg Gly Gly His Val Val Gln Lys Ala Gln Phe Asn Trp Asp 100 105 Pro Glu Thr Val Gly Leu Ile His Gly Ser Phe Phe Trp Gly Tyr Ile 120 125 Val Thr Gln Ile Pro Gly Gly Phe Ile Cys Gln Lys Phe Ala Ala Asn 135 140 Arg Val Phe Gly Phe Ala Ile Val Ala Thr Ser Thr Leu Asn Met Leu 150 155 Ile Pro Ser Ala Ala Arg Val His Tyr Gly Cys Val Ile Phe Val Arg 165 170 Ile Leu Gln Gly Leu Val Glu Gly Val Thr Tyr Pro Ala Cys His Gly 180 185 Ile Trp Ser Lys Trp Ala Pro Pro Leu Glu Arg Ser Arg Leu Ala Thr 200 205 Thr Ala Phe Cys Gly Ser Tyr Ala Gly Ala Val Val Ala Met Pro Leu 215 220 Ala Gly Val Leu Val Gln Tyr Ser Gly Trp Ser Ser Val Phe Tyr Val 230 235 Tyr Gly Ser Phe Gly Ile Phe Trp Tyr Leu Phe Trp Leu Leu Val Ser 245 250 Tyr Glu Ser Pro Ala Leu His Pro Ser Ile Ser Glu Glu Glu Arg Lys 260 265 Tyr Ile Glu Asp Ala Ile Gly Glu Ser Ala Lys Leu Met Asn Pro Val 280 Thr Lys Phe Asn Thr Pro Trp Arg Arg Phe Phe Thr Ser Met Pro Val 295 290 300 Tyr Ala Ile Ile Val Ala Asn Phe Cys Arg Ser Trp Thr Phe Tyr Leu 310 315 Leu Leu Ile Ser Gln Pro Ala Tyr Phe Glu Glu Val Phe Gly Phe Glu 325 330 Ile Ser Lys Val Gly Leu Val Ser Ala Leu Pro His Leu Val Met Thr 345 Ile Ile Val Pro Ile Gly Gly Gln Ile Ala Asp Phe Leu Arg Ser Arg 360 His Ile Met Ser Thr Thr Asn Val Arg Lys Leu Met Asn Cys Gly Gly 375 380 Phe Gly Met Glu Ala Thr Leu Leu Val Val Gly Tyr Ser His Ser





385					390					395					400
Lys	Gly	Val	Ala	Ile 405	Ser	Phe	Leu	Val	Leu 410	Ala	Val	Gly	Phe	Ser 415	Gly
Phe	Ala	Ile	Ser 420	Gly	Phe	Asn	Val	Asn 425	His	Leu	Asp	Ile	Ala 430	Pro	Arg
Tyr	Ala	Ser 435	Ile	Leu	Met	Gly	Ile 440	Ser	Asn	Gly	Val	Gly 445	Thr	Leu	Ser
Gly	Met 450	Val	Сув	Pro	Ile	Ile 455	Val	Gly	Ala	Met	Thr 460	Lys	His	Lys	Thr
Arg 465	Glu	Glu	Trp	Gln	Tyr 470	Val	Phe	Leu	Ile	Ala 475	Ser	Leu	Val	His	Tyr 480
Gly	Gly	Val	Ile	Phe 485	Tyr	Gly	Val	Phe	Ala 490	Ser	Gly	Glu	Lys	Gln 495	Pro
Trp	Ala	Glu	Pro 500	Glu	Glu	Met	Ser	Glu 505	Glu	Lys	Cys	Gly	Phe 510	Val	Gly
His	Asp	Gln 515	Leu	Ala	Gly	Ser	Asp 520	Glu	Ser	Glu	Met	Glu 525	Asp	Glu	Val
Glu	Pro 530	Pro	Gly	Ala	Pro	Pro 535	Ala	Pro	Pro	Pro	Ser 540	Tyr	Gly	Ala	Thr
His 545	Ser	Thr	Val	Gln	Pro 550	Pro	Arg	Pro	Pro	Pro 555	Pro	Val	Arg	Asp	Tyr 560

(2) INFORMATION FOR SEQ ID NO:5:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 272 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:

- (A) LIBRARY: BRAITUT02
- (B) CLONE: 754412

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

ATTTATATCA	ATGTCTTAGA	TATTGCTCCA	AGGTATTCCA	GTTTTCTCAT	GGGAGCATCA	60
AGAGGATTTT	CGAGCATAGC	ACCTGTCATT	GTACCCACTG	TCAGTGGATT	TCTTCTTAGT	120
CAGGACCCTG	AGTTTGGGTG	GAGGAATGTC	${\tt TTCTTCTTGC}$	TGTTTGCCGT	TAACCTGTTA	180
GGACTACTCT	TCTACCTCAT	ATTTGGAGAA	GCAGATGTCC	AAGAATGGGC	TAAAGAGAGA	240
AAACTCACTC	GTTTATGAAG	TTATCCCACC	TT			272

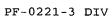
(2) INFORMATION FOR SEQ ID NO:6:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 25 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (vii) IMMEDIATE SOURCE:
 - (B) CLONE: XLR
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

CTTGATGCTC CCATGAGAAA ACTGG

(2) INFORMATION FOR SEQ ID NO:7:

25





(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 25 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:

(B) CLONE: XLF

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

AGGATTTTCG AGCATAGCAC CTGTC

25